Forecasting neutrino masses from combining KATRIN and the CMB: Frequentist and Bayesian analyses

Ole Host

Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen,
Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark

Ofer Lahav, Filipe B. Abdalla

Department of Physics and Astronomy, University College London, Gower St., London WC1E 6BT, UK

Klaus Eitel

Forschungszentrum Karlsruhe, Institut für Kernphysik, Postfach 3640, 76021 Karlsruhe, Germany
(Dated: September 7, 2007)

We present a showcase for deriving bounds on the neutrino masses from laboratory experiments and cosmological observations. We compare the frequentist and Bayesian bounds on the effective electron neutrino mass m_β which the KATRIN neutrino mass experiment is expected to obtain, using both an analytical likelihood function and Monte Carlo simulations of KATRIN. Assuming a uniform prior in m_β, we find that a null result yields an upper bound of about 0.17 eV at 90% confidence in the Bayesian analysis, to be compared with the frequentist KATRIN reference value of 0.20 eV. This is a significant difference when judged relative to the systematic and statistical uncertainties of the experiment. On the other hand, an input $m_\beta = 0.35$ eV, which is the KATRIN 5σ detection threshold, would be detected at virtually the same level. Finally, we combine the simulated KATRIN results with cosmological data in the form of present (post-WMAP) and future (simulated Planck) observations. If an input of $m_\beta = 0.2$ eV is assumed in our simulations, KATRIN alone excludes a zero neutrino mass at 2.2σ. Adding Planck data increases the probability of detection to a median 2.7σ. The analysis highlights the importance of combining cosmological and laboratory data on an equal footing.

PACS numbers: 14.60.Pq, 02.50.Tt, 95.85.Ry

One of the most important recent discoveries in physics is that neutrinos have non-zero masses. Since neutrinos are explicitly massless in the Standard Model, massive neutrinos necessarily involve new physics and may point the way for other developments. Therefore there is much need for improved knowledge of neutrino masses from both laboratory experiments and cosmological observations.

The fact that neutrinos are massive is demonstrated by atmospheric [1, 2], solar [3, 4, 5, 6, 7], reactor [8, 9, 10] and accelerator [11, 12, 13] neutrino oscillation experiments. Additionally, the masses are constrained from above by tritium beta decay [14, 15], neutrinoless double beta decay [16, 17, 18, 19], and from bounds on the amount of hot dark matter in the Universe (see below for references). Since the oscillation experiments are only sensitive to square mass differences, the absolute neutrino mass scale has not been determined, apart from the disputed observation of neutrinoless double beta decay in the Heidelberg-Moscow experiment [20, 21, 22]. However, recent cosmological observations of the Cosmic Microwave Background [23, 24, 25, 26, 27], the Large Scale Structure of the Universe [28, 29], the Baryon Acoustic Oscillations in the matter power spectrum [30], Type Ia Supernovae [31, 32, 33], and Lyman-α clouds [34] show substantial evidence for the 'standard' ΛCDM cosmological model. In that model case, the upper bound on the sum of neutrino masses $\sum m_\nu$ is driven well into the sub-eV region [27, 35, 36, 37, 38, 39, 40]. If so, it seems unlikely that a detection of the electron neutrino mass m_β will be made at KATRIN [41], the next generation tritium beta decay experiment which has a discovery potential of 0.35 eV. However, there are two important issues with the cosmological data. Firstly, the results depend on which cosmological model is used, and different models contain different numbers of parameters. Adding more parameters will in general relax the bounds on any single parameter. Secondly, it should be noted that most of the predicting power relies on Lyman-α data and it is extremely hard to model these systems [40]. This makes it challenging to assess the systematic effects associated with this measurement. Without Lyman-α, current limits are in the region 0.6–0.7 eV for $\sum m_\nu$. Hence a cleaner probe such as laboratory experiments, galaxy surveys or weak lensing surveys is very desirable. It should be noted that the next generation of large-scale surveys should be able to access the range down to 0.2 eV for the sum of the neutrino masses [42].

Comparison of various neutrino mass bounds is complicated by the fact that there are several different, but interdependent, observables involved (see, e.g., Fogli et al. [36]), but also by the use of different statistical methods which potentially leads to inconsistencies. The cosmology community usually adheres to Bayesian inference while experimental neutrino physicists tend to employ frequentist methods in oscillation experiments and abso-

lute mass searches.

In this paper we address these complications by comparing frequentist and Bayesian neutrino mass bounds in relation to the near-future KATRIN experiment and by combining the estimated KATRIN results with present and future cosmological observations in a consistent manner. First, we introduce the KATRIN experiment and then we compare the frequentist and Bayesian confidence intervals obtained from an approximate KATRIN likelihood function. Then we simulate KATRIN using Monte Carlo methods and compare a Bayesian analysis of our simulations with the frequentist analysis by the KATRIN collaboration. Finally, we predict the future neutrino mass sensitivity by combining our results with present (post-WMAP) and future (Planck) cosmological observations in a consistent Bayesian framework.

I. THE KATRIN EXPERIMENT

The Karlsruhe Tritium Neutrino experiment KATRIN is a next generation single beta decay experiment investigating the kinematics of the tritium beta spectrum near its kinematical endpoint [41]. It will measure the effective electron neutrino mass through the nuclear beta decay of tritium

$$^3\text{H} \rightarrow {}^3\text{He}^+ + \text{e}^- + \bar{\nu}_e. \qquad (1)$$

with an endpoint energy $E_0 \simeq 18.6\,\text{keV}$. Any non-zero (effective) mass of the electron antineutrino would lead to an end of the beta spectrum lower than the endpoint energy E_0 as well as to a distortion of the spectrum in the region near the endpoint.

KATRIN investigates the β-electron energy with an electrostatic high-pass filter which measures the integrated spectrum of β-electrons near the endpoint of the spectrum. The lower bound of the filter is determined by a retarding potential U and the number of electrons in a given potential energy bin is [43]

$$N(qU) = A \int_0^{E_0} \text{d}E \, \frac{\text{d}N}{\text{d}E}(E, E_0, m_\beta^2) \, f_{\text{res}}(qU, E) + N_b, \qquad (2)$$

where f_{res} is the experimental response function, E_0 is the endpoint energy of the β-spectrum, N_b is the background, and A is the signal amplitude of the spectrum. The differential electron spectrum close to the endpoint is, approximately,

$$\frac{\text{d}N}{\text{d}E} \simeq C \, (E_0 - E)\sqrt{(E_0 - E)^2 - m_\beta^2}. \qquad (3)$$

The observable is therefore the square of the effective mass. The aim of the KATRIN experiment is a sensitivity of $m_\beta < 0.2$ eV at 90% confidence in case of a null result or a 5σ discovery potential for $m_\beta \geq 0.35$ eV [43]. It has been shown [44] that non-SM contributions to the beta decay such as right-handed currents have little impact on the neutrino mass measurement.

KATRIN is actually being built at the site of Forschungszentrum Karlsruhe, having major components such as the main spectrometer vessel with its diameter of 10m and a length of 24m already installed. Commissioning of the 70m-long full setup comprising the windowless gaseous tritium source, the electron transport section, a tandem spectrometer and a semiconductor detector array is expected for 2010. To reach the expected sensitivity, a measuring time of 3 full-beam-years is anticipated, starting end of 2010.

Here, our interest in KATRIN is mainly its expected statistical performance. The benchmark is the frequentist Monte Carlo analysis by the KATRIN collaboration as given in the Design Report [43]: The outcome is an expected distribution of occurrence which is very close to Gaussian in m_β^2 with a width of $\sigma_{\text{stat.}}(m_\beta^2) = 0.018$ eV^2. This is combined with the estimate of the total systematic uncertainty, $\sigma_{\text{sys.}}(m_\beta^2) \leq 0.017$ eV^2, which leads to the total uncertainty of $\sigma_{\text{tot.}}(m_\beta^2) = (\sigma_{\text{stat.}}^2 + \sigma_{\text{sys.}}^2)^{1/2} = 0.025$ eV^2. Taking the Gaussian shape of the frequency distribution, the 90% confidence sensitivity on m_β is then derived via

$$L(90\%) = \sqrt{1.64\sigma_{\text{tot.}}}, \qquad (4)$$

and the detection threshold is $m_\beta = [5\sigma_{\text{tot.}}(m_\beta^2)]^{1/2} = 0.35$ eV.

II. BAYESIAN VS. FREQUENTIST CONFIDENCE INTERVALS

Next we discuss the frequentist and Bayesian approaches to parameter inference and calculation of upper and lower bounds.

Both frequentist confidence intervals and Bayesian credible intervals are constructed from a likelihood function $L(x, s)$ which is a function of the data x and the parameter to be inferred s (For simplicity, we consider only one parameter in this discussion). A frequentist confidence interval is a statement about frequency of occurrence. The interval $[s_1, s_2]$ is a member of the set of all intervals that fulfil the condition that the probability

$$P(s \in [s_1, s_2]) = 1 - \alpha, \qquad (5)$$

for all s, in particular the unknown true value. Here, $1-\alpha$ is the stated confidence level of the interval. The values s_1 and s_2 are determined by the likelihood function. This defining property known as 'coverage' ensures that, in the limit of a large ensemble of intervals, a fraction $1 - \alpha$ of the ensemble will contain the true value.

In the unified approach of Feldman and Cousins [45], intervals are constructed from a likelihood ordering method which involves two steps. Firstly, for a fixed choice of s a physically allowed value of x is admitted to an 'acceptance region' if the likelihood at that value is greater than at any other value not included in the interval already. In the continuous case, this means that the

acceptance region defined by $[x_1, x_2]$ will satisfy

$$\int_{x_1}^{x_2} L(x,s)\mathrm{d}x = 1 - \alpha, \qquad (6)$$

so that, additionally, the likelihood function L is everywhere greater inside the region than outside. Secondly, this procedure is repeated for every s on a finely spaced grid. Having measured the best-fit value x_0, the confidence interval is the union of all the values of s that include x_0 in their acceptance region.

On the other hand, the Bayesian credible interval is a statement about degree of belief. It attaches a probability to a range of values, indicating quantitatively how much it is believed that the actual value is within that range. Following Bayes' theorem, the posterior probability density in the true value of the parameter s is

$$P(s|x) = \frac{L(x|s)\pi(s)}{P(x)}, \qquad (7)$$

where $L(x|s)$ is the likelihood function, which is now viewed as the conditional probability of the data given s, and $\pi(s)$ is the assumed prior probability density of the parameter. The normalizing factor $P(x)$, also known as the Bayesian evidence, is simply the integral of the numerator. The credible interval $[s_1, s_2]$ is calculated by integrating the posterior probability density so that

$$\frac{1}{P(x)} \int_{s_1}^{s_2} L(x|s)\pi(s)\mathrm{d}s = 1 - \alpha. \qquad (8)$$

The framework can straightforwardly be generalized to multiple parameters, including nuisance parameters which are easy to handle. One can just integrate (marginalize) the multidimensional posterior distribution over the nuisance parameters to obtain the posterior in the parameter(s) of interest. The prior is a means of incorporating known information about s, which is then updated by the measurements in the form of the likelihood. One can think of at least three bases for the choice of prior: a known experimental result, a theoretical analysis, or 'ignorance' about the value of the parameter.

III. COMPARISON OF INTERVALS FOR KATRIN

Now we study the frequentist and Bayesian intervals constructed from the predicted KATRIN likelihood function for a vanishing mass. As discussed, the observable is the square of the electron neutrino mass m_β^2 and for the purposes of fitting it is necessary to also consider negative values of this parameter, i.e. $m_\beta^2 < 0$. The realness is then imposed afterwards. Therefore we assume a likelihood which is Gaussian in m_β^2 and has width $\sigma_{\mathrm{tot.}}(m_\beta^2) = 0.025$ eV^2, the estimated total error of KATRIN [43]. We consider varying best-fit values $m_{\mathrm{b.f.}}^2$ in the range $\pm 5\sigma = \pm 0.125$ eV^2 around zero. We calculate the frequentist confidence intervals in the unified
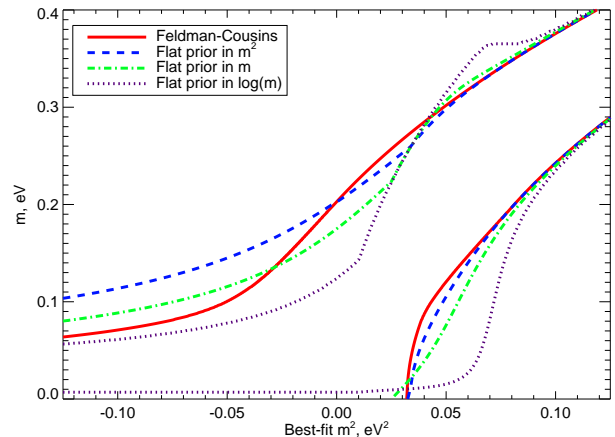

FIG. 1: Upper and lower bounds at 90% confidence as a function of the best-fit value calculated in the frequentist Feldman-Cousins approach, and in the Bayesian approach with three different priors. The interpretation of a null result, $m_{\mathrm{b.f.}}^2 = 0$ is very sensitive to the choice of analysis, while the confidence interval for a frequentist 5σ detection at $m_{\mathrm{b.f.}}^2 = 0.125$ eV^2 is virtually the same in all cases. A Gaussian likelihood function in m_β^2 is assumed. Note that, for the purpose of fitting, the observable m_β^2 can take on both positive and negative values.

approach of Feldman & Cousins [45] but we state the bounds in m_β, the parameter of interest, rather than in m_β^2. We also calculate Bayesian credible intervals for three different priors which are constant in m_β^2, m_β, and $\log m_\beta$, respectively. The first two priors are bounded from below at zero mass and the logarithmic prior is bounded at 7 meV, suggestive of lower limits from oscillation data. All the priors are bounded from above at 1 eV.

There is a certain freedom in calculating the Bayesian intervals, since an interval containing a specific probability mass is not uniquely determined without further constraints. We choose to use the shortest possible interval, integrated in the metric in which the chosen prior is constant. However, we stress that this choice is not unique.

The resulting upper and lower bounds at 90% confidence are shown in Fig. 1. The intervals, regardless of method and assumed prior, approximately converge when the best-fit value is physical and far from the boundary, i.e. for $m_{\mathrm{b.f.}}^2 \gtrsim 0.125$ eV$^2 = 5\sigma$. The behaviour for $m_{\mathrm{b.f.}}^2 \lesssim 0$ varies considerably; in particular the Feldman-Cousins upper bound drops faster than any Bayesian upper bound when going from positive to negative $m_{\mathrm{b.f.}}^2$. Also, the Bayesian upper bounds, except for the logarithmic prior, are more conservative for very negative best-fits. The behaviour of the logarithmic prior is distinct due to the larger prior probability density attached to small values of m_β. This causes the logarithmic prior to yield the tightest upper bounds for small $m_{\mathrm{b.f.}}^2$. At first glance the logarithmic prior, or Jeffreys'

prior [46] as it is also known, may appear artificial but it has the nice property that it treats all powers of the parameter equally. Also, as shown by Jaynes [47], it is the prior which contains the least information about a scale parameter, i.e. an 'ignorance' prior.

Finally, we note that the question of whether a nonzero neutrino mass is detected can also be discussed from a model selection perspective. In the frequentist framework this is the rejection of the null hypothesis that the mass is zero, while in the Bayesian sense one may compare the Bayesian evidence for models with and without the neutrino mass [48].

IV. BAYESIAN BOUNDS FROM SIMULATIONS OF KATRIN

In order to understand better the statistical performance of KATRIN from a Bayesian point of view we perform an independent Monte Carlo simulation of the experiment which is analyzed in the Bayesian framework and compared with reference results.

The model is essentially given by eqs. (2) and (3). There are four input parameters: m_β^2, the endpoint energy of the beta spectrum E_0, the background per bin N_b, which is assumed to be independent of U, and the signal amplitude of the spectrum, A. We adjust A to reproduce the statistical error of the frequentist KATRIN simulations. We also recover the approximately Gaussian shape of the likelihood function.

We assume priors which are uniform in each of the nuisance parameters and in m_β. For each realization of the data we calculate the likelihood function on a grid in the four dimensional parameter space which extends more than five standard deviations to each side of the input values. The grid contains 50 points along each of the three nuisance parameters and 130 points along the m_β^2 direction in the range $[-0.1, 1]$ eV^2 and it is constructed so that it is regular in m_β. From this tabulated likelihood function we marginalize over the three nuisance parameters and convolve the result with a Gaussian with width $\sigma_{\rm sys}(m_\beta^2) = 0.017$ eV^2 to model the maximum expected systematic error [43]. The result is the posterior probability density of m_β from which the Bayesian credible interval can be calculated as in section II.

We have checked that the resolution of the grid is sufficient by calculating the total probability contained within a given interval for ten datasets resolved on our grid and on a reference grid with 100 points for each nuisance parameter. The results differ by $\mathcal{O}(10^{-5})$.

We consider two cases with input masses $m_\beta = 0$ and $m_\beta = 0.35$ eV, respectively. For the zero input we calculate the upper bound and for the nonzero input the central value of the confidence interval, both at 90% confidence. The resulting distributions are shown in Figs. 2 and 3, respectively. Also shown are the expected upper bounds/central values (dotted lines) obtained in section II as well as the frequentist expectation.
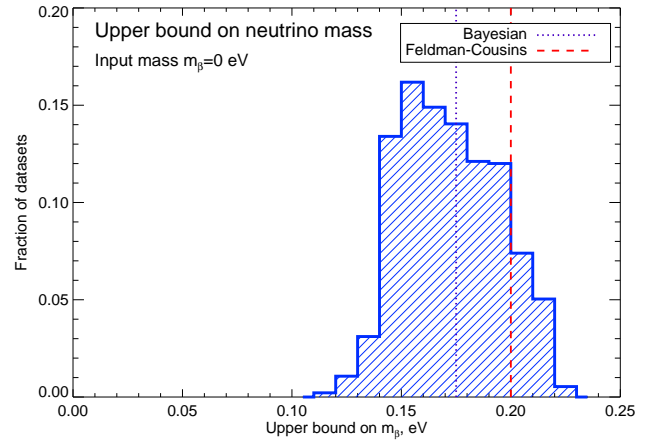

FIG. 2: Distribution of 90% upper bounds on m_β for an input $m_\beta = 0$, obtained in the Bayesian analysis assuming a uniform prior in m_β. As expected, the distribution is scattered around the upper limit from Fig. 1 for a Gaussian likelihood with a uniform prior in m_β, which is significantly lower than the frequentist expectation. Out of 1000 simulated datasets, 67 did not include zero at 90% confidence.
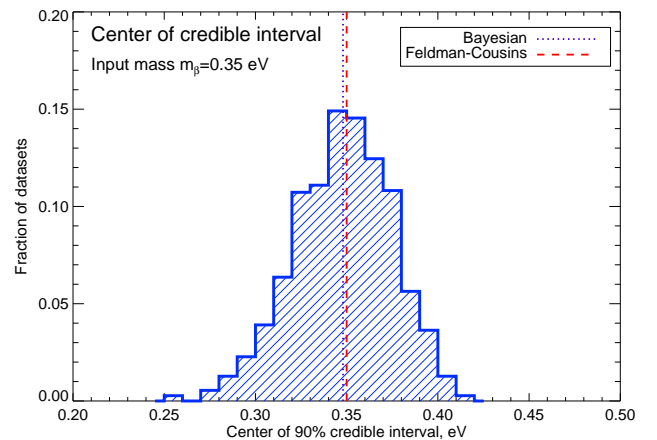

FIG. 3: Distribution of central values of 90% confidence intervals for an input $m_\beta = 0.35$ eV. As in Fig. 2, the distribution is scattered around the central value for the corresponding interval for a uniform prior in Fig. 1 but in this case there is no discrepancy with the frequentist expectation for the central value.

The mean upper bound of the distribution in Fig. 2 is 0.17 eV so it is clear that the Bayesian analysis is more optimistic than the frequentist which yields the bound 0.20 eV. This is in agreement with the analysis in section II, in particular with the upper bound of 0.175 eV obtained for a uniform prior in m_β and best-fit $m_{\rm b.f.}^2 = 0$, as shown in Fig. 1. This difference between the frequentist and Bayesian upper bound is large since, if the frequentist analysis were to yield an upper bound of 0.17 eV, the total error of $\sigma_{\rm tot.}(m_\beta^2)$ would need to be reduced by 40%.

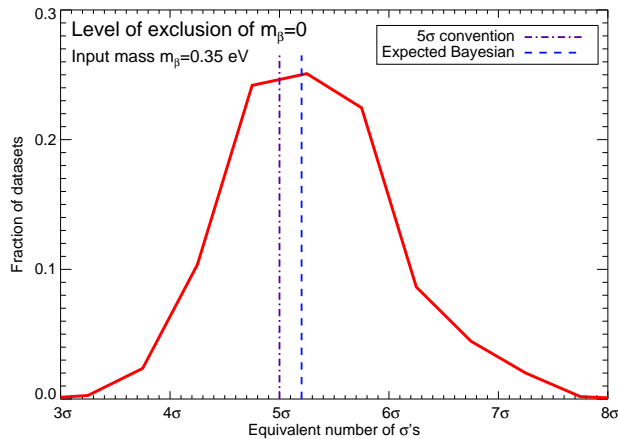FIG. 4: The level at which $m_\beta = 0$ is just excluded, given in terms of the equivalent number of σ's for the input $m_\beta = 0.35$ eV. The left vertical line is the conventional 5σ detection threshold and the 63% of the datasets to the right of this line exclude $m_\beta = 0$ at more than 5σ. The right vertical line is the expected exclusion level in the Bayesian analysis, assuming a uniform prior in m_β and a Gaussian likelihood in m_β^2. 51% of the datasets are to the right of this line.

For the input $m_\beta = 0.35$ eV in Fig. 3, the central value distribution is centered on the input and the Bayesian and frequentist analyses yield the same result. Again, this is in agreement with the intervals in Fig. 1 since the input mass is much greater than zero. The crucial question for the discovery potential, however, is at which confidence level the simulated datasets just exclude $m_\beta = 0$, in terms of the equivalent number of σ's. This is shown in Fig. 4 in terms of the equivalent number of σ's. Also indicated is the conventional 5σ discovery limit. It can be seen that a majority of the datasets exclude $m_\beta = 0$ at a greater confidence level than 5σ. The reason for this is mainly the choice of prior; a uniform prior in m_β^2 instead would only show random scatter around the 5σ line. When assuming a likelihood function which is Gaussian in m_β^2 and adopting a uniform prior in m_β, $m_\beta = 0$ is excluded at a probability corresponding to 5.2σ. This value is very close to the median of the distribution in Fig. 4.

From a frequentist point of view Fig. 4 indicates that the Bayesian credible intervals with uniform prior in m_β are more likely to undercover. This means that less than the fraction $1 - \alpha$ of an ensemble of intervals will contain the true value of the parameter—at least at this extreme confidence level.

In summary, the Bayesian analysis carried out here is more optimistic than the frequentist in this respect in the sense that both the sensitivity and discovery potential are lower.
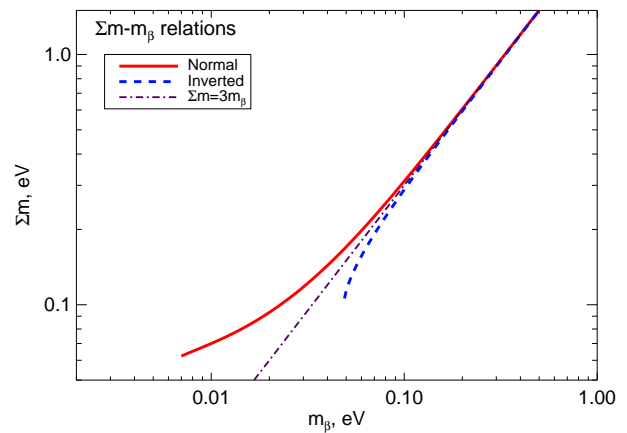

FIG. 5: Relation between m_β and $\sum m_\nu$ for normal and inverted hierarchies, and the naive relation $\sum m_\nu = 3m_\beta$. For $m_\beta \gtrsim 0.1$ eV there is virtually no differences between the relations.

V. COMBINATION WITH COSMOLOGY

Having analyzed the differences between mass bounds derived by the different statistical methods, we proceed by combining our KATRIN simulation with present and future cosmological observations in the Bayesian framework.

The observable in tritium beta decay m_β and the cosmological observable $\sum m_\nu$, are related through the neutrino mixing angles and square mass differences which are probed by oscillation experiments. The mixing angles form a parametrization of the PMNS matrix which relates the weak interaction neutrino eigenstates $|\nu_\alpha\rangle$ to the mass eigenstates $|\nu_i\rangle$ through $|\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle$ [49]. We simplify the analysis by fixing the oscillation data [8, 9, 11] at best-fit values. In particular, we assume $\sin\theta_{13} = 0$ which determines $m_\beta = (m_1 + m_2)/2$, and the masses can then be written in terms of observables as

$$m_1^2 = m_\beta^2 - \frac{1}{2}\Delta m_{12}^2, \tag{9a}$$

$$m_2^2 = m_\beta^2 + \frac{1}{2}\Delta m_{12}^2, \tag{9b}$$

$$m_3^2 = m_\beta^2 + \frac{1}{2}\Delta m_{12}^2 \pm \Delta m_{23}^2. \tag{9c}$$

The only freedom left is the hierarchy, i.e. the sign of Δm_{23}^2 in (9c). The best-fit square mass differences are

$$\Delta m_{12}^2 = 7.9 \times 10^{-5} \text{ eV}^2, \tag{10a}$$
$$\Delta m_{23}^2 = 2.6 \times 10^{-3} \text{ eV}^2. \tag{10b}$$

The hierarchy determines if the mass m_3 is greater or lesser than the pair m_1 and m_2. Relations between m_β and $\sum m_\nu$ are shown in Fig. 5.

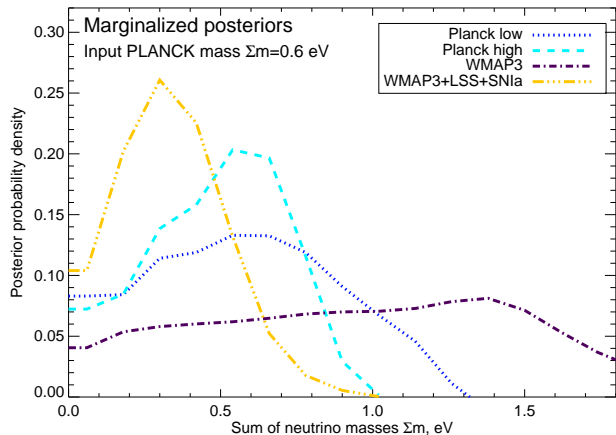In principle, the uncertainties on the neutrino mixing angles and square mass differences should be taken into

FIG. 6: The one-dimensional marginalized posteriors for $\sum m_\nu$ as obtained in in the simulation of Planck, computed with an input mass of input $\sum m_\nu = 0.6$ eV, and the WMAP analyses. These posteriors yield upper bounds of $\sum m_\nu < 1.1$ eV (Planck low), $\sum m_\nu < 0.81$ eV (Planck high), $\sum m_\nu < 1.8$ eV (WMAP only), and $\sum m_\nu < 0.66$ eV (WMAP+LSS+SNIa), all at 95% confidence.

FIG. 7: Exclusion levels of $m_\beta = 0$ given in the equivalent number of σ's for an input $m_\beta = 0.2$ eV to the KATRIN simulation, with and without cosmological data. The distribution for KATRIN alone differs from Fig. 4 only through the input. The extra data from cosmology clearly improve the detection threshold, although for this input mass a 4σ detection is still unlikely.

account, but for the purpose of this showcase we deem the best-fit values sufficient. It was found in ref. [36] that the uncertainty in the m_β–$\sum m_\nu$ relation is very small, except for the smallest allowable masses.

We combine simulated KATRIN posteriors with present cosmological data in the form of the WMAP 3-year data to which data from large-scale structure surveys and type Ia supernovae are added. Additionally, we also combine the KATRIN results with a simulation of the Planck satellite, the next cosmic microwave background satellite.

A. Bounds from the CMB

We include present limits on $\sum m_\nu$ from cosmology in the form of the analysis of the WMAP 3-year data, as performed by the WMAP team [27]. We use their MCMC chains, available at [50], for a ΛCDM cosmology described by the six minimal parameters plus the neutrino energy density. We use the chains for the largest combined dataset, which includes WMAP and other CMB data as well as large scale structure and supernovae measurements. It was shown in ref. [39] that an identified problem with the WMAP likelihood code does not affect neutrino mass bounds from combined datasets. On its own, this data set yields an upper bound on the sum of the neutrino masses of $\sum m_\nu < 0.66$ eV at 95% confidence. This strong bound is mainly caused by the large-scale structure and small-scale CMB data included; the WMAP data alone yield the much weaker constraint $\sum m_\nu < 1.8$ eV, also at 95% confidence [27, 37].

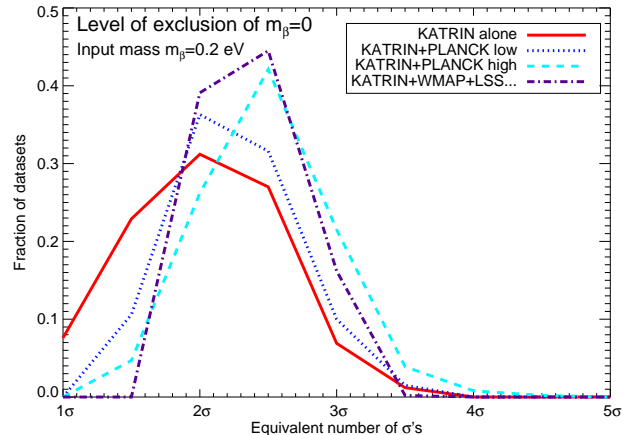Additionally, we have obtained forecasts from the

Planck satellite by using the technique described in Abdalla & Rawlings [42]. This consists mainly in assuming a fiducial model and calculating the expected rejection probability of a model around this fiducial value. We have assumed conservative values for the sensitivity by taking only one usable science channel with 8 detectors, a noise effective temperature of 120 $\mu K\sqrt{s}$, an angular resolution of 10 arcseconds, and 60% sky coverage in a one year survey. We have also considered a more optimistic forecast which would cover around 70% of the sky in a two year survey. In both cases the input fiducial mass was taken as $\sum m_\nu = 0.6$ eV and the other cosmological parameters considered were: $\Omega_m = 0.3$, $\Omega_b = 0.04$, $n_s = 0.95$, $\Omega_\Lambda = 0.7$, $n_{\rm run} = 0.00$, $\sigma_8 = 0.8$, $\tau = 0.09$, and we assumed a flat Universe. The resulting posteriors yield upper bounds of $\sum m_\nu < 1.1$ eV (conservative) and $\sum m_\nu < 0.81$ eV (optimistic) at 95% confidence.

Both the WMAP+LSS+SNIa posterior and the simulated Planck posteriors, marginalized over all parameters except $\sum m_\nu$, are shown in figure 6. The Planck simulation yields higher upper bounds on $\sum m_\nu$ since it is based on the CMB alone, and because the input to the simulation is already close to the upper bound derived from WMAP+LSS+SN.

B. Combined KATRIN and CMB

Now we combine our simulation of KATRIN with the cosmological observations discussed above.

We simulate a new set of 1000 KATRIN experiments with an input mass of $m_\beta = 0.2$ eV, the expected KATRIN 90% upper bound on a null result. Using the rela-

tions in (9), we convert each simulated m_β-posterior into a posterior in $\sum m_\nu$, assuming normal hierarchy for definiteness. For simplicity, we do not propagate the small errors from the oscillation data. The simulated KATRIN posteriors are combined with each of the cosmology posteriors individually and for each of these we calculate the confidence level at which a 'vanishing' neutrino mass is excluded. There is a technical point here since converting the KATRIN observable using (9) implies that neutrinos are massive. In other words, whereas it is possible to examine a KATRIN posterior and a cosmology posterior in isolation and determine whether there is evidence for a nonzero neutrino mass in each posterior, the two can only be combined under the assumption that both m_β and $\sum m_\nu$ are nonzero. Therefore we search for the confidence level at which the lower bound of the interval just reaches the lower limit of $\sum m_\nu$ and regard this as the detection threshold (for that posterior) of a massive neutrino. Otherwise the detection of the nonzero mass is from the oscillation data rather than the KATRIN+cosmology posterior.

The procedure yields distributions of the exclusion level of a 'vanishing' mass for KATRIN alone and for the two types of combined posteriors, which are shown in Fig. 7. For the KATRIN posteriors alone, the median detection level is 2.2σ and it is clear that a large fraction of the simulated experiments do not exclude $m_\beta = 0$ even at 2σ. On the other hand, most experiments do exclude $m_\beta = 0$ at 90% confidence, in agreement with expectations based on the analysis in section II.

The addition of the cosmological data improves the detectability of a nonzero mass. The combination with WMAP+LSS+SNIa yields a median detection level of 2.6σ while the combination the Planck yields 2.4σ (conservative) and 2.7σ (optimistic). More importantly, the widths of the distributions become smaller so that many more combined posteriors exclude a vanishing mass at more than 2σ. On the other hand, there are no detections at more than 4σ which is a consequence of the fact that neither of the three datasets are able to provide significant detections on their own. The combination of the high specs-Planck simulation with KATRIN actually performs slightly better than WMAP+LSS+SNIa and KATRIN, which is likely due to the fact that the WMAP+LSS+SNIa posterior is peaked at a much lower value of $\sum m_\nu$ than the input 0.6 eV used in the simulations. In summary, the combination of KATRIN and cosmological data may well extend the discovery potential of the experiments into regions not accessible by either type of experiment alone. We note that additional large-scale structure data could improve the detection potential further.

VI. CONCLUSIONS

We have presented a showcase for estimating the neutrino masses by combining data from laboratory and cosmological experiments. Specifically, we combine simulated results from KATRIN with cosmological probes. A key issue in this context is the consistent application of statistical methods to all data. We analyze the differing results obtained by the frequentist and Bayesian approaches, taking simulated KATRIN likelihoods as an example. Then we combine the KATRIN results with both the present cosmological constraints after WMAP and simulated future data from the Planck satellite.

The comparison of frequentist and Bayesian methods shows important differences for KATRIN, where the likelihood is likely peaked close to the physical boundary $m_\beta = 0$. While the KATRIN reference upper bound to a null result is 0.2 eV at 90% confidence, the Bayesian result is 0.17 eV, assuming a uniform prior, and even lower if a logarithmic prior is assumed. As explained in section IV, this is a sizable difference judged against the systematic and statistical uncertainties. Far away from the zero mass boundary, the different approaches converge.

The combined constraints from laboratory and cosmological data are investigated by assuming an electron neutrino mass of 0.2 eV. In this case KATRIN alone will provide a 2.2σ detection, based on 10^3 simulations of KATRIN. The 10^3 exclusion levels span a rather wide range, from zero detection to about 4σ. The addition of CMB data increases the detection level somewhat to 2.6σ (WMAP+LSS+SNIa) or $2.4\sigma - 2.7\sigma$ (Planck) but, importantly, narrows the range of exclusion levels so that in most cases a $> 2\sigma$ detection is found.

Our analysis highlights the impact of the choice of statistical analysis, and it demonstrates the application of Bayesian methods to combined datasets.

[1] Y. Ashie et al. (Super-Kamiokande), Phys. Rev. Lett. **93**, 101801 (2004), hep-ex/0404034.
[2] M. C. Sanchez et al. (Soudan 2), Phys. Rev. **D68**, 113004 (2003), hep-ex/0307069.
[3] B. Aharmim et al. (SNO), Phys. Rev. **C72**, 055502 (2005), nucl-ex/0502021.
[4] J. N. Abdurashitov et al. (SAGE), J. Exp. Theor. Phys. **95**, 181 (2002), astro-ph/0204245.
[5] J. N. Bahcall, M. H. Pinsonneault, and S. Basu, Astrophys. J. **555**, 990 (2001), astro-ph/0010346.
[6] R. Davis, Prog. Part. Nucl. Phys. **32**, 13 (1994).
[7] M. Altmann et al. (GNO), Phys. Lett. **B616**, 174 (2005),

hep-ex/0504037.

[8] M. Apollonio et al., Eur. Phys. J. **C27**, 331 (2003), hep-ex/0301017.

[9] T. Araki et al. (KamLAND), Phys. Rev. Lett. **94**, 081801 (2005), hep-ex/0406035.

[10] F. Boehm et al., Phys. Rev. **D64**, 112001 (2001), hep-ex/0107009.

[11] D. G. Michael et al. (MINOS), Phys. Rev. Lett. **97**, 191801 (2006), hep-ex/0607088.

[12] M. H. Ahn et al. (K2K), Phys. Rev. Lett. **90**, 041801 (2003), hep-ex/0212007.

[13] E. Aliu et al. (K2K), Phys. Rev. Lett. **94**, 081802 (2005), hep-ex/0411038.

[14] V. M. Lobashev, Nucl. Phys. **A719**, 153 (2003).

[15] C. Kraus et al., Eur. Phys. J. **C40**, 447 (2005), hep-ex/0412056.

[16] C. E. Aalseth et al., Phys. Atom. Nucl. **63**, 1225 (2000).

[17] C. Arnaboldi et al., Phys. Rev. Lett. **95**, 142501 (2005), hep-ex/0501034.

[18] R. Arnold et al. (NEMO), Phys. Rev. Lett. **95**, 182302 (2005), hep-ex/0507083.

[19] H. V. Klapdor-Kleingrothaus et al., Eur. Phys. J. **A12**, 147 (2001), hep-ph/0103062.

[20] H. V. Klapdor-Kleingrothaus, I. V. Krivosheina, A. Dietz, and O. Chkvorets, Phys. Lett. **B586**, 198 (2004), hep-ph/0404088.

[21] C. E. Aalseth et al., Mod. Phys. Lett. **A17**, 1475 (2002), hep-ex/0202018.

[22] H. V. Klapdor-Kleingrothaus and I. V. Krivosheina, Mod. Phys. Lett. **A21**, 1547 (2006).

[23] C.-l. Kuo et al. (ACBAR), Astrophys. J. **600**, 32 (2004), astro-ph/0212289.

[24] E. M. Leitch et al., Astrophys. J. **624**, 10 (2005), astro-ph/0409357.

[25] C. J. MacTavish et al., Astrophys. J. **647**, 799 (2006), astro-ph/0507503.

[26] J. L. Sievers et al. (2005), astro-ph/0509203.

[27] D. N. Spergel et al. (WMAP), Astrophys. J. Suppl. **170**, 377 (2007), astro-ph/0603449.

[28] S. Cole et al. (The 2dFGRS), Mon. Not. Roy. Astron. Soc. **362**, 505 (2005), astro-ph/0501174.

[29] M. Tegmark et al. (SDSS), Astrophys. J. **606**, 702 (2004),

[30] D. J. Eisenstein et al. (SDSS), Astrophys. J. **633**, 560 (2005), astro-ph/0501171.

[31] P. Astier et al., Astron. Astrophys. **447**, 31 (2006), astro-ph/0510447.

[32] A. G. Riess et al. (2006), astro-ph/0611572.

[33] W. M. Wood-Vasey et al. (2007), astro-ph/0701041.

[34] P. McDonald et al., Astrophys. J. Suppl. **163**, 80 (2006), astro-ph/0405013.

[35] O. Elgaroy and O. Lahav, Phys. Scripta **127**, 105 (2006), hep-ph/0606007.

[36] G. L. Fogli et al., Phys. Rev. **D75**, 053001 (2007), hep-ph/0608060.

[37] M. Fukugita, K. Ichikawa, M. Kawasaki, and O. Lahav, Phys. Rev. **D74**, 027302 (2006), astro-ph/0605362.

[38] A. Goobar, S. Hannestad, E. Mortsell, and H. Tu, JCAP **0606**, 019 (2006), astro-ph/0602155.

[39] J. R. Kristiansen, O. Elgaroy, and H. Dahle (2006), astro-ph/0611761.

[40] U. Seljak, A. Slosar, and P. McDonald, JCAP **0610**, 014 (2006), astro-ph/0604335.

[41] G. Drexlin (KATRIN), Nucl. Phys. Proc. Suppl. **145**, 263 (2005).

[42] F. B. Abdalla and S. Rawlings (2007), astro-ph/0702314.

[43] J. Angrik et al. (KATRIN), FZK Scientific Report 7090 (2005), URL http://www-ik.fzk.de/~katrin/publications/.

[44] J. Bonn, K. Eitel, F. Gluck, D. Sevilla-Sanchez, and N. Titov (2007), arXiv:0704.3930 [hep-ph].

[45] G. J. Feldman and R. D. Cousins, Phys. Rev. **D57**, 3873 (1998), physics/9711021.

[46] H. Jeffreys, Proc. Roy. Soc. **138**, 48 (1932).

[47] E. Jaynes, *Probability theory: the logic of science* (Cambridge University Press, 2003).

[48] R. Trotta (2005), astro-ph/0504022.

[49] W.-M. Yao, C. Amsler, D. Asner, R. Barnett, J. Beringer, P. Burchat, C. Carone, C. Caso, O. Dahl, G. D'Ambrosio, et al., Journal of Physics G **33**, 1+ (2006), URL http://pdg.lbl.gov.

[50] URL http://lambda.gsfc.nasa.gov.